

## **The Effects of Target Word Properties on the Incidental Acquisition of Vocabulary Through Reading**

**Barry Lee Reynolds**

University of Macau, Macau SAR, China

<BarryReynolds@umac.mo>

### **Abstract**

The primary aim of this investigation was to determine what combination of target word variables (frequency, patternedness, length, cognateness, lexicalization) could best predict the difficulty of incidentally acquiring vocabulary through reading. A group of adult English First Language (EL1) ( $n = 20$ ) and adult English as a Foreign Language (EFL) ( $n = 32$ ) learners were given a copy of an unmodified English novel containing nonce words to read within two weeks. After reading, they were given an unexpected meaning recall translation assessment measuring acquisition of 49 target nonce words. Results indicated that previously unknown frequently occurring shorter length cognates incidentally encountered are more likely to be acquired and retained by adult EFL learners and it was generally easier for EL1 learners to acquire and retain more frequently occurring lexicalized cognates. Results also indicate that the frequency of exposure to target words mattered more to the EFL and less to the EL1 learners. Correlations between patternedness and assessment results for both groups were not significant, further indicating no statistically significant effect of repeated surrounding contexts on incidental acquisition. Pedagogical and methodological implications are discussed.

### **Introduction**

Within language acquisition research there exists a substantial body of literature supporting extensive reading as a means of vocabulary growth for both first (L1) and second (L2) language learners (for reviews see Huckin & Coady, 1999; Krashen, 2003; Nagy & Herman, 1987; Swanborn & de Glosper, 1999; Waring & Nation, 2004). This is due in part to extensive reading being considered an excellent source of input for vocabulary that must be acquired to become advanced in the target language but which appears less frequently in aural input. Moreover, vocabulary acquisition through extensive reading has been considered as occurring incidentally because learners are focused on the task of reading instead of learning vocabulary (Hulstijn, 2001). Due to the ease of access to extensive reading materials as a source of input for learners, incidental

vocabulary acquisition has become an issue of interest in input-oriented language acquisition theories (Jenkins, Stein, & Wysocki, 1984; McQuillan, 1996; Meara, 1997; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Pigada & Schmitt, 2006; inter alia).

In recent years, incidental vocabulary acquisition has continued to gain attention, with a large number of extensive reading studies investigating the effect variables have on the incidental acquisition of vocabulary through reading. Some of the variables have included text genre (Shokouhi & Maniati, 2009), vocabulary word class (Kweon & Kim, 2008), context (Webb, 2008a), collocation (Wible, Liu, & Tsao, 2011), text modification (Negari & Rouhi, 2012), and word frequency (Hulstijn, Hollander, & Greidanus, 1996). Although a review of the incidental vocabulary acquisition literature shows evidence that a number of these variables affect the incidental acquisition of vocabulary through reading, examining these studies individually reveals many of the studies examined the effect of an isolated variable or its combined effect with frequency of exposure on acquisition without taking into consideration the possible effects of other moderating variables. Therefore, the core aim of the current study was to not only investigate whether a previously unexplored variable may affect the incidental acquisition of vocabulary through reading but in doing so also take into consideration the effects of previously investigated variables. More specifically, the following research questions are addressed:

1. What combination of investigated variables (frequency, patternedness, length, cognateness, lexicalization) predicts the difficulty of incidentally acquiring vocabulary through reading?
2. Is the incidental acquisition of vocabulary encountered while reading affected by patternedness of the context surrounding target words?

### ***Literature Review***

#### **Vocabulary Acquisition and Extensive Reading**

Although researchers seem reluctant in pointing out a single variable responsible for L1 vocabulary growth exhibited in children, several researchers have claimed that one variable in particular, direct instruction, cannot be the main reason for this growth. Generally, their conclusions are that direct instruction is very powerful in that it provides learners with a lot of explication regarding a single word's meaning and use, but when speaking in terms of breadth versus depth, learning from direct instruction is still not enough to account for the amount of vocabulary children acquire. Instead, researchers advocate a method of through context as a preferred means for learners to acquire vocabulary knowledge. Most researchers agree with Jenkins et al. (1984) that L1 vocabulary growth in children is likely due to incidental acquisition through reading.

Nagy and Herman (1987) stress that the long-term benefits of learning vocabulary from context should be taken into consideration by both language learners and educators. They claim that direct instruction is only effective for a small number of words and that the long-term effect of learning vocabulary from context encountered over lengthier periods of time outweighs the benefits of direct instruction. In fact, school-aged English as a first language (EL1) learners receive direct instruction for only a few hundred

words a year (Nagy, 1997). This means that teachers are not spending very much time on teaching vocabulary, leaving Jenkins et al. (1984) as well as Nagy et al. (1985) to draw the conclusion that most vocabulary learning occurs incidentally through reading. For example, even a modest amount (25 minutes) of reading a day for 200 days out of a year would expose EL1 learners to between 15,000 and 30,000 unfamiliar words (Nagy & Herman, 1987). If one in 20 of those words were learned, that would yield a yearly increase of between 750 and 1,500 words. Teaching the meanings of individual words that appear within student texts may only help with comprehension of those particular texts, leaving students unable to cope with other texts. Nagy and Anderson (1984) advocate wide reading as a more useful approach that could lead to multiple encounters of the same words in a variety of meaningful contexts. They further stress that “trying to expand [EL1] children’s vocabularies by teaching them words one by one, ten by ten, or even hundred by hundred would appear to be an exercise of futility” (p. 328).

Likewise, some L2 researchers claim learning incidentally is not only one strategy to acquiring a large vocabulary, it is the preferred strategy. Several scholars point out that vocabulary learning cannot happen in isolation. They suggest that when vocabulary is presented in context, the learner is expected to understand not only the meaning of the vocabulary item but also how the meaning is incorporated into the surrounding text. In this way, L2 learners get a richer sense of the word’s use and meaning (Huckin & Coady, 1999). Contexts can also provide an abundance of knowledge about target vocabulary that traditional teaching methods may not always be able to provide, including knowledge of grammatical features, collocates, situations for use, and fine aspects of meaning (Nation, 2013). Moreover, the importance of presenting vocabulary in context becomes even more obvious in terms of comprehension. Knowing how a vocabulary word contributes to the overall meaning of a text cannot be gained simply from a dictionary definition. This requires context rich exposure with an opportunity to incorporate this exposure into background knowledge.

One of the most enjoyable and contextually rich types of incidental exposure to vocabulary occurs through the reading of novels (Krashen, 1994). Horst, Cobb, and Meara (1998) called for researchers to use longer texts (i.e., chapter books, graded readers) in incidental vocabulary acquisition studies. However, besides one EL1 study (Saragi, Nation, & Meister, 1978) conducted prior to Horst et al. (1998), only four published studies have used novel-length target texts (Lehmann, 2007; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Waring & Takaki, 2003). Lehmann (2007) allowed EFL learners ten weeks to read a 235-page novel. Pigada and Schmitt (2006) allowed a French as a foreign language learner to read approximately 30,000 tokens over four weeks and Pellicer-Sánchez and Schmitt (2010) allowed an approximately 67,000-token novel to be read by EFL learners over four weeks. The time given for participants to finish reading longer target texts in previous studies was from as little as three days for EL1 readers to ten weeks for EFL readers. The studies that required participants to read within a controlled environment used two of the shortest novels (Horst et al., 1998; Waring & Takaki, 2003); the studies that allowed learners to take novels outside the classroom used relatively longer novels (Lehmann, 2007; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Saragi et al., 1978). The methodologies of these studies seem to indicate that when longer texts are selected as

target texts for participants, more naturalistic extensive reading habits are fostered whereas shorter texts are given to participants to read in more controlled settings. Giving participants target texts to finish in class would limit the length of the texts since they may not be able to finish long ones within limited class time (Pitts, White, & Krashen, 1989). Furthermore, requiring research participants to remain in a controlled environment until they finish reading does not reflect realistic language learning.

### **Incidental Vocabulary Acquisition**

Incidental vocabulary acquisition occurs when learners engaged in reading acquire vocabulary as a by-product (Nagy, 1997). Hulstijn (2001) further defines incidental vocabulary acquisition in terms of experimental research design. When participants are not aware they will receive assessment on vocabulary knowledge, any words acquired during the reading are assumed to have been acquired incidentally; expecting a future assessment would increase the likelihood of intentional vocabulary learning. Similarly, Swanborn and de Glopper (1999) define incidental vocabulary learning as the opposite of intentional vocabulary learning, where the “word incidental implies that the purpose for reading does not specifically provoke learning or directing attention to the meaning of unknown words” (p. 262).

### **Context**

One of the important variables that have been of constant interest to the incidental vocabulary acquisition research community has been the effect that contextual support has on acquisition. Researchers have investigated the claim by Saragi et al. (1978) that context can cause a difference in vocabulary acquisition. Researchers have rated the contexts in which target vocabulary appeared by how much information they believed was provided to allow learners to infer the meaning of target vocabulary. Then they tried to determine if higher rated contexts were more influential on acquisition than raw frequency of exposure. Webb (2008a), for example, exposed EFL learners to different rated one to two sentence long contexts containing target words. He found participants exposed to a few high rated contexts significantly outperformed participants exposed to more low rated contexts. This result was in contrast to previous research that tended to indicate that higher frequency of exposure would lead to greater acquisition outcomes (Horst et al., 1998; Rott, 1999; Saragi et al., 1978). According to Webb (2008a), “the quality of the context rather than the number of encounters with target words may have a great effect on gaining knowledge of meaning” (p. 232).

Other studies, however, found that additional variables were more influential than context. Zahar, Cobb, and Spada (2001), for example, came to a much different conclusion regarding the acquisition of English as a Second Language (ESL) learners than that of Webb (2007, 2008a) with EFL learners. When comparing the influence of frequency of exposure and context, Zahar et al. (2001) found “the effect of contextual support...to be subordinate to frequency” (p. 555). Furthermore, Tekmen and Daloğlu (2006), replicating Zahar et al. (2001), also found frequency to play an influential role in incidental learning of vocabulary, but it did not necessarily play a greater role for lower level learners than higher-level learners.

The seemingly fluctuating effect of frequency between studies could be due to the differences in the texts read or the properties of the words assessed, a point that has often been overlooked in previous incidental vocabulary acquisition research, making comparisons between studies troublesome. It should also be noted that other variables besides context or frequency, not of interest in these previous studies, could have also influenced acquisition outcomes.

### **Cognates**

One variable that could have affected the acquisition results reported in previous research is the cognate status of the target words assessed. Cognates are identified as words that are similar in sound and meaning between two languages (Lado, 1955). For example, in Spanish, there are numerous words that sound like certain English words and have the same meaning as the corresponding English words; therefore, Spaniards feel it easy to recognize English cognate words in their beginning stage of learning (Lado, 1955). Moreover, due to the similarity in sound and meaning, cognates facilitate English instruction as well (Lado, 1955). Daulton (2003) found eight EFL learners at a Japanese university felt cognates served as a useful guide for them to acquire English because they are easy to grasp during the learning process; in addition, these learners preferred using cognates rather than non-cognates in their written output. Similarly, Willis and Ohashi (2012) found 69 Japanese EFL learners learned and retained cognates more easily than non-cognates. Tonzar, Lotto, and Job (2009) found Italian learners were able to grasp the meaning of German and English cognates with less difficulty than non-cognates. They further found cognates act as a strong cue when translating from the L2 to the L1. Nevertheless, L1 words with the same form but different meaning in the L2 (i.e., deceptive cognates) may hinder rather than encourage L2 acquisition (Lado, 1955; Tonzar et al., 2009). Therefore, using cognates correctly requires more practice, as learners should assure whether the meaning fits the context (Daulton, 2003).

There has been little direct research on the incidental acquisition of cognates through reading. Vidal (2011), in comparing the effects of reading and listening on incidental vocabulary acquisition, did find that extensive reading facilitated the acquisition of L2 words that were cognate with L1 words. Still, L2 learners may not be aware of the L2 cognate or they may not feel confident that the L2 cognate is not a false cognate. In short, teachers will feel cognates are easier to teach and learners will feel cognates are easier to learn, that is, as long as they are aware of deceptive cognates.

### **Lexicalization**

Another factor that could make a word easier to acquire in an L2 is whether the word is lexicalized in the L1. Paribakht (2005) found L1 Farsi-speaking EFL learners knew fewer English words that were non-lexicalized than lexicalized in the EFL learners' L1 and relied more on inferencing when encountering non-lexicalized words through reading. Based on Webb (2007), Chen and Truscott (2010) conducted a study with L1 Chinese-speaking EFL learners to determine the effect of L1 lexicalization on the acquisition of seven aspects of word meaning through the reading of 1-2 sentence long contexts containing target English words. Results largely supported results reported in Webb (2007) as well as showing that words in the L2 with no lexicalized counterpart in the L1

were more difficult to learn than L2 words that had a lexicalized counterpart in the L1. The study was further replicated by Heidari-Shahreza and Tavakoli (2012) with L1-Persian speaking EFL learners; they found a significant difference in acquisition between non-lexicalized and lexicalized words.

### **Word Length**

Although it is generally accepted that longer L2 words are more difficult to acquire than shorter words, the incidental vocabulary acquisition research has not given the issue much attention. Instead, researchers have tended to ensure the length of the target words selected to appear on the assessments is fairly equal. Laufer (1990), however, does provide a review of pivotal studies investigating the effect of word length on the acquisition of L2 vocabulary that indicate longer words are generally more difficult to acquire than shorter words but this significance seems to disappear as the proficiency in the L2 improves.

### **Patternedness**

Although the lack of investigation on particular variables could be somewhat responsible for the difference in acquisition outcomes of previous research, differences in the effects of certain variables such as context investigated by Webb (2007) and Zahar et al. (2001) could also be due to the type of context in which the target words appeared. In Webb (2007) EFL learners read isolated sentences whereas in Zahar et al. (2001) ESL learners read a story taken from their ESL textbooks. This distinction highlights the importance of operationalizing similar terminology used in describing research variables. Furthermore, this opens the research community up to exploring other types of contextual support. The current investigation aims to determine whether the occurrence of target words within repeated surrounding context affects their incidental acquisition. Here patternedness refers to a reoccurring n-gram surrounding a target word in the target text read by participants. Incidental L2 vocabulary acquisition research has emphasized an importance in investigating the number of contextual exposures learners require in order to acquire a word, however, in doing so research has not taken into consideration an enriched contextual view of language learning. Specifically, incidental vocabulary acquisition research has not investigated the effect repetition of target words in the same surrounding context may have on acquisition, with only a few studies (e.g., Pigada & Schmitt, 2006; Rott, 1999) mentioning the need for research to factor in the effects of collocations and other phraseological units on the incidental acquisition of vocabulary through reading. Specifically, would the incidental acquisition of a word encountered n-times in the same contiguous sequence of words be similar to a word encountered the same number of times but in different sequences of words? Would the repeated surrounding context somehow help to induce the incidental acquisition of the target word? Much of the past research on incidental vocabulary acquisition has examined the frequency variable by counting typographical units separated by white space on a page and looking for a correlation between frequencies and acquisition, failing to consider different types of contextual support. The current investigation is the first to have considered whether a target word appearing in the same surrounding context (i.e., the same word forms surrounding a target word) could have an effect on incidental vocabulary acquisition.

It seems appropriate here to reference the famous quote by Firth (1957, p. 11): “You shall know a word by the company it keeps.” Without considering the influence of the surrounding context of a target word, very little can be known about the acquisition of the word in question. Moon (1997, p. 43) cautions against results gained from considering words in isolation because as soon as words are part of a text they are “meaningful and inseparable units” with the other co-occurring words appearing within the text. Research by scholars such as Bolinger (1976, 1977) and Pawley and Syder (1983) stress that formulaic sequences are widespread in language and people are dependent on memorization of such sequences for faster processing. Taking a contextual view of vocabulary through the consideration of target words’ repeated appearance within the same context could provide more insight into the incidental acquisition of vocabulary through reading.

Previous research has shown similarities and differences in how L1 and L2 speakers of a language process multiword patterns. Ullman (2005) asserts that when a person processes native language vocabulary, then two memory systems are employed: a declarative memory system is used for the storage of memorized phrases and a procedural system is used for processing rules of a language. Ullman (2005) further theorizes that when the same person is processing an L2, especially in early stages of L2 development, the declarative memory system is relied on more than the procedural memory system. Wray (2002) notes that in the beginning stages of learning an L2 the use of formulaic sequences allow for an easy way to express meaning and to communicate (e.g., Hi, how are you?; Fine, thank you.; I don’t know.; where is the [noun]?). However, as L2 learners develop the ability to create novel sentences, their ability to use formulaic language seems to fall short and they produce utterances that are markedly non-native like (e.g., \*in the other hand; \*in my point of view; \*pay time on). One reason for this difficulty in L2 learners to acquire a native-like use of formulaic language could be due to their inability to notice these chunks, especially if they only encounter them through text. Wible (2008) draws to attention the fact that text-oriented EFL learners, unlike native speakers, are often first exposed to language as text and somehow must discover that the word forms separated by white space on the page fit together to form language chunks. EFL learners do not have the advantage of being able to acquire the aural language chunk first to contrast it with the written form. This may be indicative that EFL learners would benefit more from having more exposure to repeated contexts.

Using self-paced reading of sentences containing multiword patterns, Kim and Kim (2012) investigated whether the frequency of collocating pairs of words that appeared within the multiword patterns would have an effect on the recognition of the multiword patterns by EL1 and ESL learners of English. Their prediction that frequency is a factor that affects the degree to which ESL learners holistically store and retrieve multiword units during processing was correct. The EL1 learners’ response times for low frequency targets were significantly longer than both mid-level and high frequency targets, but the ESL learners only showed a difference between the low and high frequency targets. It appears as if EL1 learners store most multiword units as chunks whereas ESL learners only store “those that are most widely used and encountered in their daily life” (p. 838).

Huang, Wible, and Chou (2012) conducted a study to investigate whether formulaic input would affect EL1 learners and EFL learners' language processing. In a pre-test they found that both EL1 and EFL learners showed a processing advantage for formulaic sequences over non-formulaic sequences; however, EL1 learners were able to recognize formulaic sequences one word position earlier than EFL learners. Taking formulaic sequences used in the pretest that showed no processing advantage for EFL learners, it was investigated whether frequency (providing more encounters) or salient input (i.e., underlining) could facilitate EFL learners' processing of the formulaic sequences. Eye tracking results showed that frequency was more effective at facilitating EFL learners to process the formulaic sequences as fixed units than underlining, yet, as with the other formulaic sequences from the pre-test, EFL learners still recognized the formulaic sequences one word position later than EL1 learners.

The results of Huang et al. (2012) support those of previous research (e.g., Conklin & Schmitt, 2008; Schmitt & Underwood, 2004) but also show how EFL learners can improve processing of formulaic language when provided adequate input. The advantage of frequency over highlighting in making formulaic language more salient to EFL learners is of interest to the incidental L2 vocabulary acquisition research community because highlighting and other types of text enhancement have been argued as being more effective than frequency of exposure (Bruton, López, & Mesa, 2011). The influence of the appearance of target vocabulary within the same repeated context (i.e., patternedness) has yet to be considered as a potential influence on the acquisition of vocabulary encountered incidentally through reading.

Reviewing the literature on formulaic language found a lack in previous research that specifically addresses incidental acquisition of target vocabulary that appears within multiword patterns. However, there are still some findings that are pertinent to the discussion of incidental acquisition of vocabulary. Past research has shown differences in the processing of multiword patterns by native and non-native speakers. Some studies have shown positive results of L2 learners acquiring the ability to detect and also incidentally acquire collocations through reading. Results also seem to indicate that with more exposure, L2 learners should become more equipped at multiword processing. However, no incidental vocabulary acquisition studies have taken a look at how the repeated appearance of unknown target words within the same surrounding context may affect incidental acquisition through reading.

The majority of L2 vocabulary research has treated vocabulary as individual words isolated by white space on a page. Schmitt (2010) points out that this is due to convenience; the isolated word as a lexical unit is convenient to identify, teach, and research. However, there is an increasing awareness that language learners do not learn, process, or produce language word by word. Instead, learners learn, produce, and process language in chunks. It is with these thoughts in mind that the current research was undertaken to determine whether the repeated exposure of unknown target words within the same surrounding context would have an effect on incidental vocabulary acquisition through reading.



## ***Methodology***

### **Research participants**

Although previous incidental vocabulary acquisition studies have been conducted for different L1s and L2s, few incidental vocabulary acquisition studies have compared acquisition results between native and non-native speakers of the same target language. To allow for more controlled comparison between studies it must be ensured as with the present investigation that participant groups have read the same target text and been assessed over the same target vocabulary. This required the recruitment of both EL1 and EFL participant groups. The focus of the current investigation was on the incidental acquisition of vocabulary by adult learners receiving tertiary education in their home countries. EFL instead of ESL learners were intentionally chosen to take part in this investigation since they receive less exposure to aural input, thus relying heavily on textual input for language learning.

**EL1 learners.** The EL1 learners needed to be monolingual English speakers and to have never read the target text prior to taking part in the investigation. To locate willing participants, a recruitment letter was given to students enrolled in two general education English literature courses and one general education introduction to applied linguistics course from two mid-western state universities in the United States. At the time of the experiment most of the EL1 learners ( $n = 20$ ; female = 16; male = 4;  $M = 33$  yr.;  $SD = 14$ ;  $Mdn = 31$ ) were studying for undergraduate degrees ( $n = 18$ ), and there was one master's and one Ph.D. student. The EL1 learners were studying for a range of degrees including: applied sciences ( $n = 5$ ); computer science ( $n = 3$ ); liberal arts ( $n = 4$ ); business & management ( $n = 7$ ); and undecided/university studies ( $n = 1$ ). Each EL1 learner was given a bookstore gift certificate for taking part in the study.

**EFL learners.** The EFL learners needed to be studying EFL at the time of the experiment and to have never read the target text prior to taking part in the investigation. An intact class of undergraduate L1 Mandarin Chinese speakers ( $n = 32$ ; female = 18; male = 14;  $M = 21$ ;  $SD = .95$ ;  $Mdn = 21$ ) enrolled in an advanced English novel reading elective course at a national university in northern Taiwan were recruited as the EFL learners to take part in the investigation. The EFL learners were majoring in applied sciences ( $n = 19$ ) and liberal arts ( $n = 13$ ). To ensure that the EFL learners vocabulary knowledge was adequate to allow for unobstructed reading of the target text, participants were administered the Vocabulary Size Test (VST) (see Nation & Beglar, 2007). The EFL learners were administered the VST after the reading task and assessments were completed. At the time of the investigation, they had received EFL instruction for 11-12 years. Since the EFL learners were reading the novels as part of their regular classroom regime, they were not given bookstore gift certificates for taking part in the study.

### **Procedures**

Before the experiment began, both the EL1 and EFL learners were given a detailed handout including directions regarding how to complete the reading task. They were unaware that their vocabulary acquisition would be assessed, because it was assumed that being made aware in advance that vocabulary would be tested could affect the attention subjects paid to vocabulary encountered in the text. Since this study seeks to

investigate the incidental acquisition of target vocabulary through reading, the participants were asked not to consult references or discuss the novel contents with others. They were specifically told "Please read the novel in your leisure time without consulting any other references (for example: the Internet, dictionaries) or discussing the novel with others (for example: face-to-face with classmates, on-line chats, posting to social networking sites, over the phone)." On the day the EL1 and EFL learners expected to be either interviewed or participate in a class discussion, they received a vocabulary assessment.

**Vocabulary size test.** Nation (2006) reports that in order for L2 learners to have unobstructed comprehension of written English they need to have acquired about 8,000-9,000 word families. The EFL learners were administered the on-line traditional Chinese version of the Vocabulary Size Test (VST) located on the Test Your Skill VocabularySize.com website (<http://my.vocabularysize.com/>) (Victoria University of Wellington, 2010). After removal of the three outliers, scores on the VST for the remaining 29 L2 Experimental participants ranged from 5,000 to 13,500 with an average score of around 9,300. The average score obtained by the participants nearly reached the results of doctoral students obtained by Nation and Beglar (2007), indicating the majority of participants would not have difficulty in reading the target text chosen for this research. The EL1 learners were not administered the VST.

**Reading task.** Each participant was given a hard copy of the 37,611-token target text to read within two weeks. [1] The EFL learners were given time to read during three classes (class 1 = 1 hour; class 2 = 3 hours; class 3 = 1 hour), but due to absences, some participants were unable to take advantage of all five class hours given for reading. If they did not finish reading in class, they were told to finish reading outside of class. The classroom teacher had already given the EFL learners the freedom to carry novels wherever they wanted and read whenever they wanted for other reading tasks; therefore, the reading task mimicked how the EFL learners usually read novels for the class.

**Target text selection.** *The BFG* (Dahl, 1982) was chosen as the target text because of the use of *giant* words. The *giant* words are made-up words used by one of the main characters in the novel, The Big Friendly Giant (i.e., The BFG) and the other giants that live near him. These *giant* words are unknown to participants reading this novel for the first time. This means that participants could read the novel outside the classroom without the possibility of exposure to the giant words in other texts they may encounter. Furthermore, there was no need to administer a pretest, again because exposure to the giant words will not occur outside the novel. Eggins (2004) describes the *giant* words used in *The BFG*:

*as conform[ing] to possible phonological combinations of English, ...exploit[ing] the phonaesthetic qualities of English sound combinations...incorporate[ing]...the grammar of English, through the attachment of conventional English morphemes of tense and word class...Thus the grammatical and phonological resources of the language function conventionally. (p. 28)*

In other words, the *giant* words used by the BFG in the novel should not present any more difficulties to the participants than any other unknown English words they might encounter while reading. Below is an excerpt from the novel in which The BFG introduced Sophie, a little orphan girl, to his favorite drink, *froboscottle*:

*'Froboscottle,' announced the BFG. 'All giants is drinking froboscottle.'*

*'Is it as nasty as your snozzcumpers?' Sophie asked.*

*'Nasty!' cried the BFG. 'Never is it nasty! Froboscottle is sweet and jumbly!'*

*He got up from his chair and went to a second huge cupboard. He opened it and took out a glass bottle that must have been six feet tall. The liquid inside it was pale green, and the bottle was half full. (p. 44)*

To determine whether the target text would allow for unobstructed reading by the L2 English-speaking experimental group, the novel was subjected to analysis using the RANGE computer program (Heatley, Nation, & Coxhead, 2002). The RANGE output for each list contains the raw number of tokens and types as well as their representative percentages; a family number is also given. The results for the novel *The BFG* show that over 86% of its tokens appear in the first 2000 word families, indicating that the majority of words that appear in the novel are high frequency words. However, results also show that 4.17% of the tokens did not register on any of the lists. Manual analysis of the RANGE results showed a majority of these "off list" tokens represented names (e.g., *BFG* = 516 tokens; *Bloodblotter* = 35 tokens); onomatopoeic words not yet included in the 16th list (e.g., *poo*, *owch*, *owelee*), and a few real English words that did not register (e.g., *helpings*, *foggiest*, *bunt*). Out of the 299 suitable word types found in the target text, 43 occurred at a frequency of three or higher; therefore, these 43 as well as an additional six randomly chosen nonce words from the remaining 256 with a frequency of two were selected as the target words for the research. See Appendix A for individual coding of target words.

**Coding for frequency and length.** Frequency of occurrence was calculated by allowing for inflectional and derivational variation in the form of target words when counting target word tokens (i.e., tokens of the exact word form as well as inflected and derived forms). Frequency ranged from 2 to 91 for the target words ( $M = 8.12$ ;  $SD = 13.97$ ;  $Mdn = 4$ ;  $MO = 3$ ). Word length was measured as the number of letters that constituted a word. Length ranged from 5 to 19 letters ( $M = 9.69$ ;  $SD = 2.80$ ;  $Mdn = 9$ ;  $MO = 9$ ).

**Coding of patternedness.** Language patterns can be analyzed to determine if they affect incidental vocabulary acquisition through reading. Wray (2002, p. 19), in a discussion of identifying formulaic language, states that "...you have to have a reliable set of representative examples, and these must therefore have been identified first." She goes on to suggest collecting "...particular linguistic material and then hunt[ing] through it in some more or less principled way, pulling out strings which, according to some criterion or group of criteria, can justifiably be held up as formulaic" (Wray, 2002, p. 20). Nattinger and DeCarrio (1992) suggest scrutinizing a six-gram (i.e., token) window adjacent to both sides of a word when searching for language patterns. Using these guidelines would allow for the analysis of target texts to determine if target words reoccur in the same context.

*Patternedness* refers to the reoccurrence of target words within the *same surrounding context* (i.e., a reoccurring n-gram surrounding a target word in the target text read by participants). Target words were coded as either *patterned* or *non-patterned* according to the percentage of a particular target word's tokens found appearing within a repeated context. A threshold of 50% was set for the proportion of tokens of a target word needed to appear within a repeated context for the target word to be coded as *patterned*. This was done by aligning the sentences containing the tokens of each target word in a similar fashion as a concordancer aligns the results from user queries. With the target word tokens in the center, the contexts surrounding the target word tokens could easily be compared for repeated contexts. Examples of the process for two target words (one coded *patterned* and one coded *non-patterned*) are exemplified in Figures 1 and 2. Coding resulted in 37 (73%) target words coded as *patterned* and 12 (17%) as *non-patterned*.

**Coding of cognateness.** Target words that met the following three requirements put for by Kondrak (2009) were coded as *cognate* or *non-cognate*: 1) reoccurring sound correspondences, 2) phonetic similarity, and 3) semantic affinity. For example, the target word *chiddlers* (i.e., children) was coded as *cognate* whereas the target word *gigglehouse* (i.e., a girl's school) was coded as *non-cognate*. After coding, 30 target (61%) words were coded as *cognate* and 19 (39%) as *non-cognate*.

**Coding of lexicalization.** Only target words that were lexicalized in English were coded as *lexicalized*; all other target words were coded as *non-lexicalized*. For example, the target word *chiddlers* (i.e., children) was coded as *lexicalized* whereas the target word *squifflerotter* (i.e., an insult of some sort) was coded as *non-lexicalized*. After coding, 41 (84%) target words were coded as *lexicalized* and 8 (16%) as *non-lexicalized*.

**Meaning recall assessment.** Previous studies (e.g., Chen & Truscott, 2010; Webb, 2007) have used translation as a means to assess research participants' recall of vocabulary meaning and form. For this research, a meaning recall assessment was constructed in which research participants were given an alphabetized list of the 49 target words with instructions to provide as much information as possible, including a direct translation, about the meaning of the target words in the blank space provided (see Figure 3). EL1 learners were given directions in English and EFL learners were given directions in Chinese. For this research, only answers that contained the correct translation or a correct definition were scored as correct. For example, for the target word *bellypoppers* (i.e., *helicopters*) the English translation of airplane and the Chinese translation of 飛機 was considered incorrect; although both airplanes and helicopters are types of machines that fly, only the answer helicopters in English or 直升機 in Chinese was scored as correct. Likewise, a definition of *an aircraft capable of hover, vertical and horizontal flight in any direction* 一台能夠任意在空中盤旋、垂直及水平飛行的飛行器 would be considered correct whereas the definition *an aircraft with wings driven by jet engines* 一台雙翼並以噴射引擎為動力的飛行器 would be considered incorrect. To further ensure intra-rater reliability, the raters were asked to read through

Patterned Example

Step 1: Listing of all target word tokens for a target word with a span of six words on either side.

1	I	is	knowing	there	is	a	gigglehouse	for	girls	and	I	is	guzzling
2				What	is	a	gigglehouse	for	girls				
3	I	is	wanting	dreams	about	a	gigglehouse	for	girls	and	boggleboxes	for	boys

Step 2: Starting from the positions next to the node word, delete words in the span that only occur once.

1					is	a	gigglehouse	for	girls	and			
2					is	a	gigglehouse	for	girls				
3						a	gigglehouse	for	girls	and			

Step 3: If repetition of word forms are different on either side of the node, delete those that appear on the side of the form of the target word occurring less frequently.  
Context length should be sacrificed for pattern frequency.

1						a	gigglehouse	for	girls				
2						a	gigglehouse	for	girls				
3						a	gigglehouse	for	girls				

Step 4: Calculate percentage of tokens appearing in repeated context.

$$3 \div 3 = 1.00$$

.50 or higher = patterned  
.49 or lower = non-patterned

Figure 1. Example of process for coding a target word as patterned.

Non-Patterned Example

Step 1: Listing of all target word tokens for a target word with a span of six words on either side.

1				You	is	having	bellypopers	is	you	not		
2					Wiz hoppy	big	bellypopers					
3	him	underneath	the	belly	of	your	bellypopers	and	carry	him	like	a periods
4	please	bring	back	here	in	the	bellypopers	all	my	collection	of	dreams
5	you	is	taking	these	clutching	only	bellypopers	any	close			
6	we	is	driving	back	to	the	bellypopers	and	picking	up	the	lightsome guests

Step 2: Starting from the positions next to the node word, delete words in the span that only occur once.

1							bellypopers					
2							bellypopers					
3							bellypopers	and				
4						the	bellypopers					
5						the	bellypopers					
6							bellypopers	and				

Step 3: If repetition of word forms are different on either side of the node, delete those that appear on the side of the form of the target word occurring less frequently.  
Context length should be sacrificed for pattern frequency.

1							bellypopers					
2							bellypopers					
3							bellypopers					
4						the	bellypopers					
5						the	bellypopers					
6							bellypopers					

Step 4: Calculate percentage of tokens appearing in repeated context.

$2 \div 6 = .33$   
.50 or higher = patterned  
.49 or lower = non-patterned

Figure 2. Example of process for coding a target word as non-patterned.

all the assessment sheets before they began coding and to constantly refer to the provided English and Chinese translations during coding. Both also had access to sentence long contexts in which the 49 target nonce words appeared. To ensure inter-rater reliability, the two raters were asked to complete the rating independently. Agreement was reached for nearly 99.8% of the answers (4,498 answers); for the remaining 10 answers, the raters met and discussed the answers until agreement was reached. The raters were bilinguals—literate in both English and Chinese, having received either an MA or PhD in Applied Linguistics. Both had experience in the teaching of Chinese/English and English/Chinese translation.

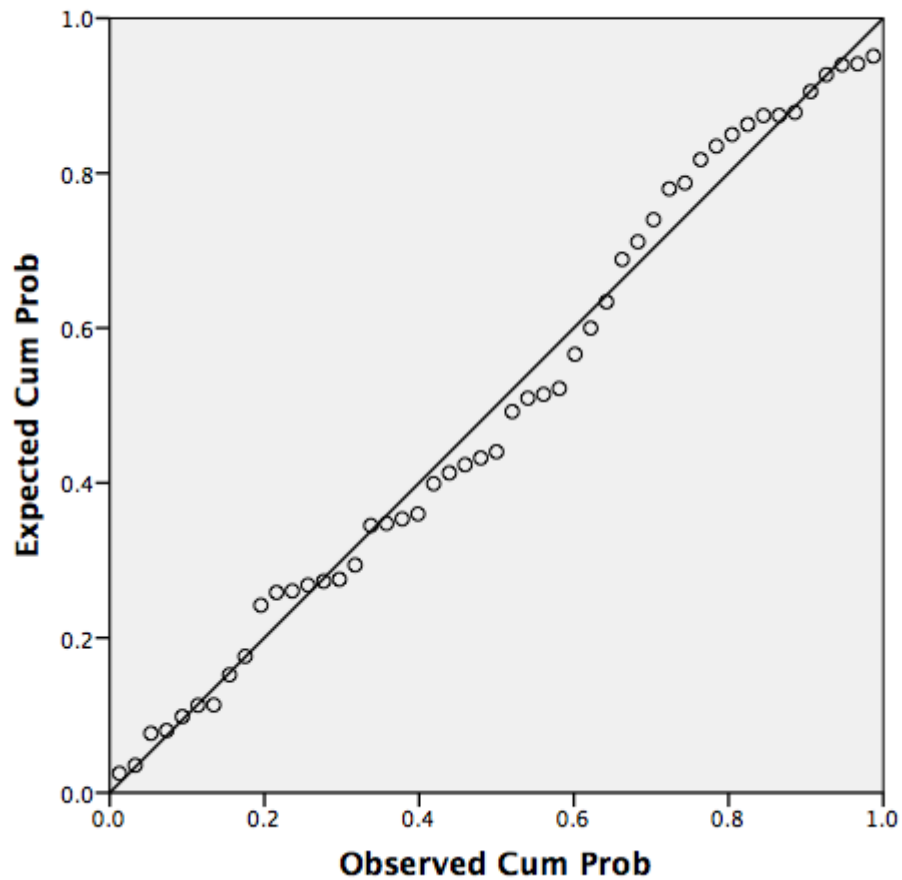
<p><b>Part 1</b></p> <p>全名:</p> <p>簡介: 用中文表達你了解以下這些字的意思 (至少要寫上中文翻譯)。</p> <p>2. bellypoppers</p>
<p><b>Part 1</b></p> <p>Full Name:</p> <p><u>Directions:</u> Tell me everything you know (including an English translation) about the meaning of the words below:</p> <p>2. bellypoppers</p>

**Figure 3. Example meaning recall assessment item.**

### **Data Analysis**

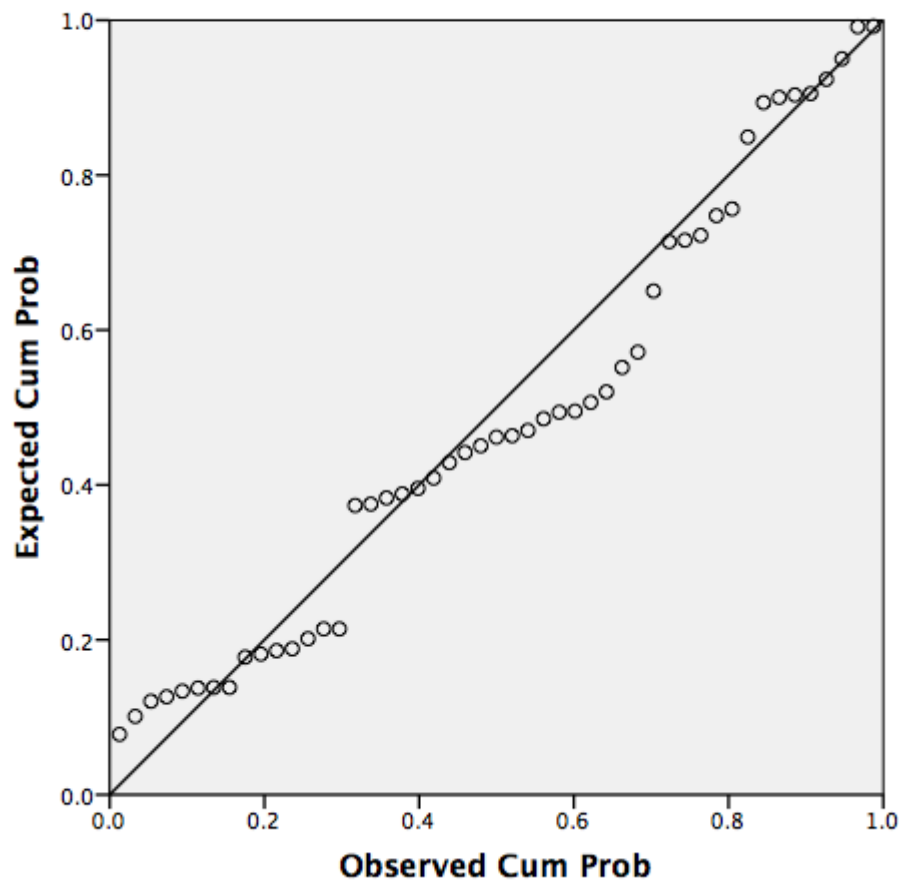
The meaning recall results were analyzed for both participant groups. In addition to Pearson correlations, multiple regression models of the data were found using the method of sequential regression. The assumptions of multiple regression were assessed. First, it was necessary to test for multicollinearity between the variables by looking for any high correlations; therefore, Pearson correlations were run for both the EL1 and EFL data. Fidell and Tabachnick (2001) caution against correlations higher than  $r = .70$ ; none of the correlations for both data sets were found to be above .70. Next, P-P plots and Cook's D were examined to determine normal distribution of the data. The P-P plot of the EL1 data showed a little curvature of the points in the distribution; this is evidence of near normality in the data set, therefore regression analysis could continue (see Figure 4). The P-P plot of the EFL data showed slight curvature of points in the distribution; this is evidence of some non-normality in the data set, but it was only slight and therefore regression analysis could continue (see Figure 5). The maximum Cook's distance value found in the EL1 data set was .296 and .261 for the EFL data set, indicating there were no outliers that are of concern. Examining the data using a

scatterplot between the studentized residuals and the predicted value of the standardized residuals show some slight signs of heteroscedasticity for the EFL data but not for the EL1 data.



**Figure 4. *P-P* plot for EL1 data.**





**Figure 5. *P-P* plot for EFL data.**

Item Facility (IF) was measured as the response variable to run the correlations and multiple regressions. “Item facility (*IF*) is defined here as the proportion of students who answered a particular item correctly” (J. D. Brown, 2003, p. 17). For example, if 10 out of 36 participants answered an item correctly, then the IF for that item would be 0.278 ( $10 \div 36 = 0.278$ ), indicating that about 28% of the participants were able to answer that item correctly. Previous research has indicated that frequency of exposure plays an influential role in the acquisition of vocabulary incidentally acquired through reading (R. Brown, Waring, & Donkaewbua, 2008; Horst, 2005; Horst et al., 1998; Hulstijn et al., 1996; Pellicer-Sánchez & Schmitt, 2010; Rott, 1999; Tekmen & Daloğlu, 2006; Webb, 2008b; Zahar et al., 2001). Since the current investigation was also particularly interested in the possible effect of Patternedness on incidental acquisition, Patternedness was entered after Frequency and prior to the other variables previously shown to affect acquisition (see Willis & Ohashi, 2012). The explanatory variables were entered in the following order: Frequency, Patternedness, Length, Cognateness, and Lexicalization (Patternedness, Cognateness, and Lexicalization were coded and entered as binary variables).

## Results

### Descriptive Statistics

EL1 item facility was higher than EFL item facility for the 49 target word items, indicating that EL1 learners acquired more target words incidentally than EFL learners. Looking at the three binary variables in terms of item facility found a similarity in the pattern for EL1 and EFL learners. EL1 item facility of patterned words was higher than non-patterned words. EFL item facility of patterned words was higher than non-patterned words. EL1 item facility of cognate words was higher than non-cognate words. EFL item facility of cognate words was higher than non-cognate words. EL1 item facility of lexicalized words was higher than non-lexicalized words. EFL item facility of lexicalized words was higher than non-lexicalized words. Descriptive statistics are displayed in Table 1.

**Table 1. Descriptive statistics for item facility displayed by variable for both learner groups**

Group	Total IF (n = 49)	Patterned (n = 36)	Non-Patterned (n = 13)	Cognate (n = 30)	Non-Cognate (n = 19)	Lexicalized (n = 41)	Non-Lexicalized (n = 8)
EL1	.518 (.281)	.538 (.281)	.462 (.285)	.592 (.297)	.401 (.213)	.518 (.286)	.313 (.131)
EFL	.213 (.288)	.237 (.308)	.147 (.220)	.314 (.325)	.054 (.084)	.251 (.300)	.020 (.033)

*Note.* Standard deviations in parentheses; n = target words.

### Correlations

Table 2 shows the Pearson correlations between Item Facility (IF) and the explanatory variables for the EL1 learners' meaning recall results. Three significant correlations were found. The explanatory variable that correlates most strongly with IF is Cognateness, followed by Lexicalization, and finally Frequency. These correlations are positive, indicating that it was generally easier for the EL1 group to learn and retain more frequently occurring lexicalized cognates.

**Table 2. Correlations between variables for EL1 Meaning Recall**

	IF	Frequency	Patternedness	Length	Cognateness	Lexicalization
IF						
Frequency	.325*					
Patternedness	.122	.229				
Length	.005	-.143	-.133			
Cognateness	.334**	.125	.186	-.315*		
Lexicalization	.326*	.128	.235	-.268*	.555***	

*Note.* \* < .05, \*\* < .01, \*\*\* < .001.

Table 3 shows the Pearson correlations between IF and the explanatory variables for the EFL learners meaning recall results. Four significant correlations were found. The explanatory variable that correlates most strongly with IF is Frequency, followed by Cognateness, then Lexicalization, and finally Length. All the correlations are positive except between IF and Length, indicating that it was generally easier for the EFL group to learn and retain shorter more frequently occurring lexicalized cognates.

**Table 3. Correlations between variables for EFL Meaning Recall**

	IF	Frequency	Patternedness	Length	Cognateness	Lexicalization
IF						
Frequency	.456***					
Patternedness	.140	.229				
Length	-.242*	-.143	-.133			
Cognateness	.444***	.125	.186	-.315*		
Lexicalization	.300*	.128	.235	-.268*	.555***	

Note. \* < .05, \*\*\* < .001.

### **Multiple Regression Models**

Two sequential regressions were run to examine the effects of the individual explanatory variables on the meaning recall scores for the EL1 and EFL learners (explanatory variables not shown to significantly correlate with IF were dropped out of the models). The EL1 multiple regression model found using the sequential procedure consisted of three explanatory variables: Frequency, Cognateness, and Lexicalization. The EFL multiple regression model found using the sequential procedure consisted of four explanatory variables: Frequency, Length, Cognateness, and Lexicalization. Table 4 shows that Model 3, with all three explanatory variables, accounted for 21.5% of the variance in meaning recall scores for the EL1 learners. Table 5 shows that Model 4, with all four explanatory variables accounted for 36.5% of the variance in the meaning recall scores for the EFL learners. According to the models, frequently occurring cognates will be more likely to be acquired and retained by both groups of learners.

**Table 4. Multiple Regression Models found using a sequential regression procedure with EL1 meaning recall data**

Model	<i>r</i>	<i>r</i> <sup>2</sup>	<i>r</i> <sup>2</sup> Change	Frequency <i>Beta</i>	Cognateness <i>Beta</i>	Lexicalization <i>Beta</i>
1	.325	.106	.106*	.325		
2	.439	.193	.087*	.297	.288	
3	.464	.215	.022	.277	.199	.180
* <i>p</i> < .05 significant <i>F</i> change between current and previous model using an ANOVA						

**Table 5. Multiple Regression Models found using a sequential regression procedure with EFL meaning recall data**

Model	<i>r</i>	<i>r</i> <sup>2</sup>	<i>r</i> <sup>2</sup> Change	Frequency <i>Beta</i>	Length <i>Beta</i>	Cognateness <i>Beta</i>	Lexicalization <i>Beta</i>	
1	.456	.208	.208***	.456				
2	.490	.240	.032	.430	-.181			
3	.603	.364	.124**	.399	-.068	.373		
4	.604	.365	.001	.398	-.065	.354	.036	
** <i>p</i> < .01, *** <i>p</i> < .001 significant <i>F</i> change between current and previous model using an ANOVA								

### ***Discussion and Implications***

The results showed that the largest contributor to acquisition (i.e., IF) for both groups as shown on the meaning recall results was frequency, followed by cognateness. Furthermore, adding the variables of length and lexicalization to the models did little in explaining the total variance of the meaning recall scores. These results have several implications for pedagogy and research methodologies used for investigating the incidental acquisition of vocabulary through reading. Below these implications are discussed.

First, the result confirms previous research showing that repeated encounters with vocabulary affects the likelihood of incidental acquisition (Pellicer-Sánchez & Schmitt, 2010; Rott, 1999; Saragi et al., 1978; Zahar et al., 2001). One of the most convenient ways for adult EFL learners to receive vocabulary exposure is through extensive reading and there have been several successful studies showing that extensive reading can lead to increased vocabulary knowledge by adult EFL learners (Daskalovska, 2011; Pellicer-Sánchez & Schmitt, 2010; Sabet, Barekat, & Fayazi, 2013; Sato, 2012); however, as Sheu (2003) points out, in Taiwan extensive reading is less accepted as a route of vocabulary development with rote methods receiving more acceptance. Even for contexts in the greater China region that claim faithfulness to aims of extensive reading methods of language education, once investigated further, it is evident that publishers and teachers are not actually adhering to those aims (Renandya, Hu, & Xiang, 2015). Still, some researchers in the Taiwan context have found extensive reading to aid learners in acquiring contextually dependent vocabulary features (Wang, 2013), for example, to acquire collocations requires repeated exposure to different occurrences since no grammar rules exist to explain why certain word pairs collocate while other word pairs do not (Webb, Newton, & Chang, 2013). In addition, the positive effect repeated exposure to words in context has on acquisition can be further enhanced if the repeated words appear within self-selected texts (Bai, 2011; Lee, 2005, 2006).

Although contextualized exposure to vocabulary can lead to a deeper understanding and eventual acquisition of words, this conclusion does not indicate the process of acquisition is efficient. In other words, extensive reading is considered as a convenient route to acquire vocabulary, but it may not always be the quickest route to increasing

one's vocabulary (Cobb, 2007); therefore, to increase acquisition efficiency, a systematic extensive reading program that allows for manipulation of frequency through the use of computer assisted language learning technologies (Cobb, 2007, 2008; Horst, 2005) or through graded readers (Nation & Wang, 1999) is recommended. Such systems provide and guarantee routes for learners to encounter frequent enough vocabulary exposure that will thus increase the likelihood of acquisition after exposure. However, the frequency effect only occurs when learners recognize repeated exposures to vocabulary as such. Post-hoc analysis of previous incidental vocabulary acquisition studies has shown researchers to operationalize repeated exposure in their studies differently (Reynolds & Wible, 2014). It is therefore necessary as has been done in the current investigation to clearly operationalize frequency of occurrence, indicating whether the counting of only exact tokens, lemmas, or word families has been used to investigate the frequency effect. This allows for more direct comparisons to other studies. In addition, extensive reading of printed or digital texts should not be the only form of lexical input learners receive, instead these texts can and should be combined and incorporated into other incidental learning (Reynolds, 2014) and interactive learning activities (Racine et al., 2013). In fact, a meta-analysis of extensive reading research by Nakanishi (2014) indicates extensive reading leads to gains in reading comprehension and should be incorporated into any language-learning curriculum. As the meta-analysis suggests, extensive reading and frequent input can also induce other aspects of language learning besides vocabulary (Ellis, 2012).

Discussing the general acquisition of vocabulary is quite straightforward, especially when considered only in conjunction with frequency of exposure. However, when certain vocabulary features are considered, it becomes more difficult to draw direct conclusions. Still, there are some guidelines that EFL teachers and researchers may want to consider. Frequency is flexible and can be manipulated, but cognateness is a fixed word property that cannot be controlled. Some have recommended teachers in Japan devote less class time to explaining the meanings of cognates, since less exposure is likely needed to encourage acquisition (Willis & Ohashi, 2012). However, recent research (Reynolds, 2016) with Taiwanese EFL learners indicates they are unaware of the quantity of English and Chinese cognates, which often results in awkward translations, especially for local cuisine (Su & Du, 2015). Teachers in Taiwan, therefore, may wish to remind students these cognates do exist, especially for concepts originating from Anglophone cultures. As for incidental vocabulary acquisition researchers, they should consider the effect of cognateness on acquisition when nonce words are used that mimic the real words they represent. Pitts et al. (1989) reported the use of nonce words to be problematic in their study. "Two items had abnormally high scores. The first item was *nochy*, meaning 'night'. This is close to the Spanish word for 'night', *noches*. The second item was *moloko*, meaning 'milk'" (p. 273). Significant correlations were found between IF and cognateness in both data sets, indicating that cognates played a facilitative role in the incidental acquisition of vocabulary through reading. For the EL1 group this is indicative that the readers may have been able to associate the nonce word with the corresponding English word. The results from the EFL group points towards the conclusion that they may have already acquired the real English word and therefore were able to associate the already acquired English words with their cognate nonce words. Future research

should investigate for this “nonce word effect” by comparing gain scores of two comparable experimental groups, one group receiving exposure to nonce words and another the real word equivalents. Future studies that wish to utilize nonce words as target words and control for the effect of cognateness should ensure that the nonce words do not share the cognate features (i.e., reoccurring sound correspondences, phonetic similarity, and semantic affinity) of participants’ L1 or L2.

Comparisons between the two groups that were involved in the current investigation should be made carefully, especially considering the different environmental conditions of the two groups. Furthermore, the EFL learners were younger than the EL1 learners. Nakanishi’s (2014) meta-analysis of extensive reading research showed the effect of extensive reading increases with age, but “...more research is required to confirm this hypothesis” (p. 21). Nakanishi failed to find a significant difference in the effect of extensive reading for university students ( $d = 1.12$ ) and adults ( $d = 1.48$ ). So, although there is some possibility that age could be responsible for some difference in the acquisition results reported in the current investigation, it is likely the difference is not significant. Any difference found might be due to the L1 language backgrounds. The correlation between frequency and IF for the EL1 learners was  $r = .325$  and  $r = .456$  for the EFL learners. A tentative interpretation is frequency mattered more to the EFL learners than the EL1 learners. In addition, there was a slight difference between the two groups in terms of the correlation between cognateness and IF, with  $r = .334$  for the EL1 learners and  $r = .444$  for the EFL learners.

Correlations between patternedness and IF for both learner groups were not found to be statistically significant. Recurrent patterns of the surrounding context do not appear to have positively affected the incidental acquisition of target words. This conclusion was drawn through the analysis of vocabulary assessment data collected after reading. However, there is no way of knowing how the participants treated the target words or the recurring surrounding contexts while reading. With the advent of eye trackers, it is possible to conduct future research that gives a better picture of what happens when L1 and L2 speakers are exposed to unknown vocabulary through reading. For instance, Williams and Morris (2004) found the context surrounding novel words received longer fixation time than the context surrounding familiar words; however, they did not look for a difference in fixation time for contexts that were repeated and those that were not repeated for novel words.

### ***Limitations and Further Study***

The mode frequency of the target words used in the current investigation is three. Since previous research has shown the more frequent the input, the more likely participants are able to recognize multiword patterns as such, it is therefore not surprising that there was not a statistical correlation between IF and patternedness. Therefore, the results of the current study are limited in that target words did not appear in more frequency recurring context. In addition to the use of eye tracking processing measurements, future incidental vocabulary acquisition through reading research should make comparisons between patterned and non-patterned vocabulary occurring at different frequencies of exposure. Using such a method to manipulate frequency will allow for a

more controlled comparison of the effects of patternedness on the incidental acquisition of vocabulary through reading.

The current research defined patternedness as reoccurring n-grams containing the target words naturally occurring in the target text given to participants to read. Wray and Perkins (2000) provide a good overview of all the different multiword pattern types (e.g., polywords, phrasal constraints, meta-messages, sentence builders, situational utterances, verbatim texts; *inter alia*) reviewed in previous literature. The current study investigated the effect of only one type of multiword pattern on the incidental acquisition of vocabulary encountered through reading. Future research should be conducted to determine whether other types of patternedness occurring in target texts positively affect the incidental acquisition of vocabulary.

Previous research has shown that the more proficient a person is in a target language, the better the person may perform at incidentally acquiring vocabulary in the language (Tekmen & Daloğlu, 2006); although the vocabulary sizes of the EFL learners was obtained, future research should try to measure other language outcomes, such as comprehension, especially when longer target texts are given to participants. Unfortunately, while trying to construct a reading task more representative of realistic reading habits some control over extraneous variables was loosened. Future research should investigate whether allowing participants the freedom to read outside of a controlled environment affects acquisition outcomes. Lastly, the current investigation measured vocabulary growth after the reading had been done. Therefore, only the acquisition of the vocabulary was measured and analyzed. Future research should attempt to complement such data with real-time processing data to provide a more comprehensive picture of what is occurring when learners encounter unknown words through reading.

The current investigation was undertaken to determine whether several variables (i.e., frequency, patternedness, length, cognateness, and lexicalization) affected the incidental acquisition of vocabulary through reading by adult EL1 and EFL learners. Results indicated that frequency and cognateness had a noticeable effect on acquisition for both learner groups with frequency mattering more to the EFL learners. The results of the investigation provide support for the claim that reading can be a viable route for L1 and L2 vocabulary development. It is hoped that the current investigation can draw attention to the need of more research simultaneously investigating for the effects of multiple variables on the incidental acquisition of vocabulary through reading.

### **Note**

Hyphens and apostrophes are treated as spaces; therefore, *I've* is counted as two tokens as is *cross-legged*.

### **Acknowledgements**

This research was partially supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant No. MOST 103-2410-H-010-015-.

## About the Author

**Barry Lee Reynolds**, Ph.D., is Assistant Professor of English Education in the Faculty of Education at the University of Macau, Macau SAR, China. After obtaining 13 years' experience in teaching ESL/EFL in the USA and Taiwan, in 2016 he relocated to the University of Macau to devote himself to TEFL teacher training and research into subfields of Applied Linguistics including L1/L2 Vocabulary Acquisition, Vocabulary Translation, L2 Literacy Instruction, Educational Technology, and Pre-Service Teacher Training. His published work has appeared in *TESOL Quarterly*, *Reading Research Quarterly*, *English Today*, *Applied Linguistics Review*, *Computers & Education*, *British Journal of Educational Technology*, among others.

## References

- Bai, Y.-L. (2011). *The degree of freedom of reader choice and second language incidental vocabulary learning*. (MA Thesis), National Central University, Chungli, Taoyuan, Taiwan.
- Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum*, 1(1), 1-14.
- Bolinger, D. (1977). Idioms have relations. *Forum Linguisticum*, 2(2), 157-169.
- Brown, J. D. (2003). Norm-referenced item analysis (item facility and item discrimination). *Shiken: JALT Testing & Evaluation SIG Newsletter*, 7(2), 16-19.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20(2), 136-163.
- Bruton, A., López, M. G., & Mesa, R. E. (2011). Incidental L2 vocabulary learning: An impracticable term? *TESOL Quarterly*, 45(4), 759-768. doi:10.5054/tq.2011.268061
- Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, 31(5), 693-713. doi:10.1093/applin/amq031
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3), 38-63.
- Cobb, T. (2008). Commentary: Response to McQuillan and Krashen (2008). *Language Learning & Technology*, 12(1), 109-114.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- Dahl, R. (1982). *The BFG*. New York: Puffin.
- Daskalovska, N. (2011). The impact of reading on three aspects of word knowledge: Spelling, meaning and collocation. *Procedia Social and Behavioral Sciences*, 15, 2334-2341. doi:10.1016/j.sbspro.2011.04.103



- Daulton, F. E. (2003). The effect of Japanese loanwords on written English production-A pilot study. *JALT Hokkaido Journal*, 7, 4-14.
- Eggins, S. (2004). *An Introduction to Systemic Functional Linguistics*. New York: Continuum.
- Ellis, N. C. (2012). What can we count in language, and what counts in language acquisition, cognition, and use? In S. T. G. D. Divjak (Ed.), *Frequency effects in language learning and processing* (pp. 7-33). Germany: Walter de Gruyten GmbH & Co.
- Fidell, L. S., & Tabachnick, B. G. (2001). *Using Multivariate Statistics* (4th ed. ed.). Boston, MA: Allyn and Bacon.
- Firth, J. R. (1957). *Papers in Linguistics, 1934-1951*. London: Oxford University Press.
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *RANGE and FREQUENCY programs* (Version 1.32). Retrieved from [http://www.victoria.ac.nz/lals/staff/Publications/paul-nation/Range\\_BNC.zip](http://www.victoria.ac.nz/lals/staff/Publications/paul-nation/Range_BNC.zip)
- Heidari-Shahreza, M. A., & Tavakoli, M. (2012). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition by Iranian EFL learners. *The Language Learning Journal, iFirst*, 1-16. doi:10.1080/09571736.2012.708051
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, 61(3), 355-382.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.
- Huang, P.-Y., Wible, D., & Chou, C.-t. (2012). EFL learners' mental processing of multi-word units. *English Teaching & Learning*, 36(1), 129-163. doi:10.6330/ETL.2012.36.1.04
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, 21, 181-193.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258-286). Cambridge: Cambridge University Press.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80(3), 327-339.
- Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, 21(4), 767-787.
- Kim, S. H., & Kim, J. H. (2012). Frequency effects in L2 multiword unit processing: Evidence from self-paced reading. *TESOL Quarterly*, 46(4), 831-841. doi:0.1002/tesq.66
- Kondrak, G. (2009). Identification of Cognates and Recurrent Sound Correspondences in Word Lists. *TAL*, 50(2), 201-235.
- Krashen, S. D. (1994). The pleasure hypothesis. In J. E. Alatis (Ed.), *Educational Linguistics, Crosscultural Communication, and Global Interdependence* (pp. 299-322): Georgetown University Press.

- Krashen, S. D. (2003). Free voluntary reading: Still a very good idea. In L. Bridges (Ed.), *Exploration in language acquisition and use* (pp. 15-29). Portsmouth: Heinemann.
- Kweon, S.-O., & Kim, H.-R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language*, 20(2), 191-215.
- Lado, R. (1955). Patterns of difficulty in vocabulary. *Language Learning*, 6(1-2), 23-41.
- Laufer, B. (1990). Why are some words more difficult than others? - Some intralexical factors that affect the learning of words. *International Review of Applied Linguistics Teaching*, 28(4), 293-307.
- Lee, S.-y. (2005). Sustained silent reading using assigned reading: Is comprehensible input enough? *The International Journal of Foreign Language Teaching*, 1(4), 10-12.
- Lee, S.-y. (2006). A one-year study of SSR: University level EFL students in Taiwan. *The International Journal of Foreign Language Teaching*, 2(1), 6-8.
- Lehmann, M. (2007). Is intentional or incidental vocabulary learning more effective? *The International Journal of Foreign Language Teaching*, 3(1), 23-28.
- McQuillan, J. (1996). How should heritage languages be taught?: The effects of a free voluntary reading program. *Foreign Language Annals*, 29(1), 56-72.
- Meara, P. (1997). Towards a new approach to modeling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40-63). Cambridge: Cambridge University Press.
- Nagy, W. E. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 64-83). Cambridge: Cambridge University Press.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19(3), 304-330.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270.
- Nagy, W. E., & Herman, P. A. (1987). Breadth and dept of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hillside, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233-253.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nation, I. S. P., & Wang, M.-t. K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, 12(2), 355-380.
- Nattinger, J. R., & DeCarrio, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Negari, G. M., & Rouhi, M. (2012). Effects of lexical modification on incidental vocabulary acquisition of Iranian EFL students. *English Language Teaching*, 5(6), 95-104. doi:10.5539/elt.v5n6p95
- Paribakht, T. S. (2005). The influence of first language lexicalization on second language lexical inferencing: A study of Farsi-speaking learners of English as a foreign language. *Language Learning*, 55, 701-748. doi:doi: 10.1111/j.0023-8333.2005.00321.x
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmitt (Eds.), *Language and communication* (pp. 192-226). London: Longman.
- Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language*, 22(1), 31-55.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1), 1-28.
- Pitts, M., White, H., & Krashen, S. D. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language*, 5(2), 271-275.
- Racine, J. P., Benevides, M., Graham-Marr, A., Coulson, D., Browne, C., Poulshock, J., & Waring, R. (2013). Vocabulary acquisition, input, and extensive reading: A conversation. *Language Teacher*, 37(4), 57-60.
- Renandya, W. A., Hu, G., & Xiang, Y. (2015). Extensive reading coursebooks in China. *RELJ Journal, online first*, 1-19. doi:10.1177/0033688215609216
- Reynolds, B. L. (2014). Evidence for the task-induced involvement construct in incidental vocabulary acquisition through digital gaming. *The Language Learning Journal, online first*, 1-19. doi:10.1080/09571736.2014.938243
- Reynolds, B. L. (2016). Troublesome English translations of Taiwanese dishes. *English Today*, 32(2), 15-23. doi:10.1017/S026607841500067X
- Reynolds, B. L., & Wible, D. (2014). Frequency in incidental vocabulary acquisition research: An undefined concept and some consequences. *TESOL Quarterly*, 48(4), 843-861. doi:10.1002/tesq.197
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21, 589-619.

- Sabet, M. K., Barekat, B., & Fayazi, S. M. (2013). Incidental vocabulary learning through extensive reading by Iranian intermediate EFL learners. *International Journal of Social Science Tomorrow*, 2(3), 1-8.
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6(2), 72-78.
- Sato, F. (2012). Vocabulary acquisition through extensive reading. *Accents Asia*, 6(1), 52-69.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. London: Palgrave Macmillan.
- Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 173-189). Amsterdam: John Benjamins Press.
- Sheu, S. P.-H. (2003). Extensive reading with EFL learners at beginning level. *TESL Reporter*, 36(2), 8-26.
- Shokouhi, H., & Maniati, M. (2009). Learners' incidental vocabulary acquisition: A case on narrative and expository texts. *English Language Teaching*, 2(1), 13-23.
- Su, T.-Y., & Du, Y.-J. (2015, November 10, 2015). 「field of ah hua」 是啥？ 大學餐廳英文連篇錯 [(field of ah hua) is what? Even a university cafeteria gets it wrong]. Retrieved from <http://news.tvbs.com.tw/life/news-625258/>
- Swanborn, M. S. L., & de Glosper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261-285.
- Tekmen, E. A. F., & Daloğlu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, 39(2), 220-243.
- Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning*, 59(3), 623-646.
- Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory and practice* (pp. 141-178). Washington DC: Georgetown University Press.
- Victoria University of Wellington. (2010). VocabularySize – Free tools to measure your students' word knowledge. Retrieved from <http://my.vocabularysize.com/>
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219-258. doi:10.1111/j.1467-9922.2010.00593.x
- Wang, Y.-H. (2013). Incidental vocabulary learning through extensive reading: A case of lower-level EFL Taiwanese learners. *The Journal of Asia TEFL*, 10(3), 59-80.
- Waring, R., & Nation, I. S. P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English-Speaking World*, 4, 11-23.

- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 130-163.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46-65. doi:10.1093/applin/aml048
- Webb, S. (2008a). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232-245.
- Webb, S. (2008b). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79-95. doi:http://dx.doi.org/10.1017/S0272263108080042
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91-120. doi:10.1111/j.1467-9922.2012.00729.x
- Wible, D. (2008). Multiword expressions and the digital turn. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 163-180). Amsterdam: John Benjamins Publishing Company.
- Wible, D., Liu, A. L.-E., & Tsao, N.-L. (2011). A browser-based approach to incidental individualization of vocabulary learning. *Journal of Computer Assisted Learning*, 27(6), 530-543. doi:10.1111/j.1365-2729.2011.00413.x
- Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, 16(1), 312-339.
- Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *The Language Learning Journal*, 40(1), 125-137. doi:10.1080/09571736.2012.658232
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20, 1-28.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, 54(4), 541-572.

## Appendix A

Nonce Word	English	Recal 1 EFL N=32	Recal 1 EL1 N=20	Frequenc y	Patternednes s	Lengt h	Cognatenes s	Lexicalizatio n
beans	being	0.94	1.00	91	1	5	1	1
bellypoppers	helicopters	0.19	0.60	6	1	12	0	1
bogthumper	nightmare	0.03	0.45	3	1	10	0	1
chiddlers	children	0.56	0.90	10	1	9	1	1
chittering	chatting	0.25	0.90	3	0	10	1	1
crocadowndillies	crocodile	0.19	0.95	3	0	16	1	1
dillions	millions	0.56	0.95	3	1	8	1	1
disgustable	disgusting	0.84	0.95	6	1	11	1	1
elefunt	elephant	0.59	1.00	4	1	7	1	1
filthing	filthy	0.03	0.95	7	1	8	1	1
flushbunking	hurrying	0	0.05	8	1	12	0	0
frightsome	frightening	0.69	0.90	8	1	10	1	1
froboscottle	soda pop	0.31	0.60	20	1	11	0	1
gigglehouse	school for girls	0.09	0.60	3	1	11	0	0
glumptious	delicious	0	0.70	3	0	10	0	1
grinksludger	insult	0	0.25	3	0	12	0	0
grobswitcher	a bad dream	0.03	0.35	6	1	12	0	0
horriddest	horrid	0.63	0.95	3	0	9	1	1
jabbeling	jabber	0	0.75	4	1	9	1	1
jumbly	rumble	0	0.10	4	1	6	1	1
majester	majesty	0.56	0.95	19	1	8	1	1
micies	mice	0.59	1.00	4	1	6	1	1
norphan	orphan	0.84	1.00	2	1	7	1	1
phizzwizard	great dream	0.03	0.55	7	1	11	0	0
poisnowse	poisonous	0.59	0.80	3	0	9	1	1
replusant	repulsive	0.06	0.80	3	1	9	1	1
ringbeller	good dream	0	0.10	2	0	10	0	0
rotsome	rotten	0.09	0.80	6	0	7	1	1
scrotty	terrible	0	0.10	4	1	7	0	1
scrumdiddlyumptious	scrumptious	0	0.95	7	1	19	1	1
scuddling	scalding, scuttle	0	0.45	4	1	9	1	1
snitching	snatching	0.09	0.45	8	1	9	1	1
snortling	snorting, snoring	0.06	0.70	3	0	9	1	1
snozzcubmers	cucumber	0.72	0.95	44	1	12	1	1

snozzling	sleeping	0.09	0.40	2	0	9	0	1
spikesticking	stabbing	0.06	0.25	3	1	13	0	1
squifflerotter	insult	0	0.25	2	0	14	0	0
squiffling	amazing	0	0.05	2	0	10	0	1
squinkers	fast moving	0	0	2	0	9	0	0
swiggle	swig, lots, wiggling	0	0.45	4	1	7	1	1
swollop	swallow	0.44	0.70	4	1	7	1	1
thingalingaling	thing	0	0.75	3	1	15	1	1
titchy	tiny	0	0.20	7	1	6	1	1
tottler	toddler	0	0.95	3	1	7	1	1
trogglehumper	nightmare	0.03	0.40	13	1	13	0	1
venomsome	venomous	0.06	0.90	3	1	9	1	1
whizzpopping	farting	0.16	0.80	19	1	12	0	1
whoppsy	wonderful	0	0.10	12	1	7	1	1
winkles	eyes	0	0.25	5	1	7	0	1

© Copyright rests with authors. Please cite TESL-EJ appropriately.